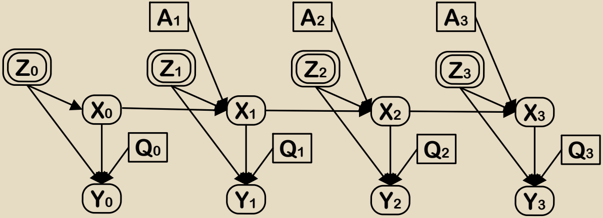
An importance sampling algorithm for cognitive diagnostic models using restricted regresion

Russell Almond

Educational Psychology and Learning Systems
College of Education
Florida State University

Florida Educational Research Association

POMDP model of educational processes



- X_t (θ_t) — Latent proficiency process
- Y_t — Observable outcomes
- Z_t — Background variables
- A_t — Action taken at each time step
- Q_t — Measurement plan (Q-matrix)

Challenges in the POMDP model

- Estimating student trajectories with known parameters (particle filter).
- *Find parameters of evidence models (single time slice).*
- Find parameters of proficiency growth model (causal model for actions).
- Find optimal measurement plan (sequence of Q's).
- Find optimal plan for student (sequence of actions conditioned on observations).

Particle Filter: Estimating student trajectories

- Sequential importance sampling:
  1. Simulate R trajectories: $X_0^{(r)}, \ldots, X_T^{(r)}$.
  2. Calculate weight $w^{(r)}$ for generating observations given trajectory.
  3. $\tilde{X}_t = \sum_r w^{(r)} X_t^{(r)}$
- Calculations factor iteratively across time slices.
- Works for arbitrary choice of (cross-sectional) evidence model and (longitudinal) proficiency growth model.

Evidence models: The cross sectional piece

The cross-sectional piece of the model often takes on a familiar functional form.

Latent Variable	Observed Variable	Model
Normal	Normal	Regression, Factor Analysis
Normal	Discrete	(Multivariate) IRT
Discrete	Normal	Conditional Gaussian, Clustering
Discrete	Discrete	Bayes net, CDMs

Often a matrix Q is used to determine which proficiency variables are relevant for which observations.

$q_{jk} = 1$ iff Proficiency k is relevant for Observable j.

BQ-Regression: Restricted regression models

- Restrict to the normal-normal case.

$$\mathbf{Y}_t = \mathbf{B}_t\mathbf{X}_t + \mathbf{b}_{0t}\mathbf{1} + \mathbf{E}_t \qquad \text{where} \quad \mathbf{E}_t \sim \mathcal{N}(\mathbf{0}, S_{\mathbf{Y}_t\mathbf{Y}_t.\mathbf{X}_t}) \ . \tag{1}$$

- Restrict $b_{jk} = 0$ if $q_{jk} = 0$.
- Call this a *BQ-Regression*
- Could be missing data in Y. (Assume MAR)
- Need to be able to weight observations (importance sampling)
- Restrict to a single time slice (drop t).

Introduction

- Goal is to find B, b_0 and $S_{\mathbf{YY}.\mathbf{X}}$ subject to restriction Q.
- The jth row can be found by regressing Y_j on the X_k values for which $q_{jk} = 1$.
- The sweep operator (Beaton, 1964; Dempster, 1969) will calculate the appropriate coefficients from the covariance matrix.
- To get BQ-regression, for each observable just sweep out X values which correspond to the 1's in that row of the Q-matrix.
- Can get residual covariance matrix by calculating residuals and then calculating sum of squares.
- More work is needed if any of the $\mathbf{Y}$ values are missing.

Missing data and the sweep operator

- Little and Rubin (1986/2002) use the Sweep operator as part of an EM algorithm for missing data in the multivariate normal setting.
- Assumes data are missing at random.
- For each missing data pattern:
  1. Sweep the matrix $\mathbf{T}$ to predict the missing values for this pattern from the observed value.
  2. Do a regression imputation for the missing value.
  3. Adjust the covariance matrix for for expected covariance (particularly, diagonal) (Let $\mathbf{T}''$ be the matrix of adjustments. Final adjsted matrix is

$$\mathbf{T}^{(i+1)} = \mathbf{Y}_{+}^{(i)T}\mathbf{W}\mathbf{Y}_{+}^{(i)} + \mathbf{T}'$$

- Converges in one pass for monotone missing data patterns.
- For non-monotone patterns requires EM algorithm.

Calculating the residual covariance matrix.

- Tricky part is calculating residual covariance matrix in the presence of missing data.
- Looks like E-step above, only now uses patterns in Q-matrix rather than missingness patterns.
- Add partial covariance matrix to covariance matrix adjustment ($\mathbf{T}'$) as before.
- Cross-product terms should be okay if *local independence assumption* (observables independent given latent variables) holds.

Importance Sampling

- Assume $\mathbf{X}$ is normally distributed with parameters $\boldsymbol{\pi}$.
- Let evidence model parameters be $\boldsymbol{\Omega}$.
- Estimate $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$ using EM-algorithm.
- E-step is

$$\int p(\mathbf{Y}|\mathbf{X},\mathbf{Q},\boldsymbol{\Omega}^{(i)})p(\mathbf{X}|\boldsymbol{\pi}^{(i)})d\mathbf{X} = \prod_{n=1}^{N} \int p(\mathbf{y}_n|\mathbf{x}_n,\mathbf{Q},\boldsymbol{\Omega}^{(i)})p(\mathbf{x}_n|\boldsymbol{\pi}^{(i)})d\mathbf{x} \tag{2}$$

- Key idea: Use Monte Carlo integration to tackle the integral.

Stochastic E-step

- For each individual n, draw R possible realizations of $\mathbf{x}_n$,

$$\mathbf{x}_n^{(1,i)}, \ldots, \mathbf{x}_n^{(R,i)} \ .$$

- Calculate weights based on likelihood of generating data sequence.

$$w_n^{(r,i)*} = p(\mathbf{y}_n | \mathbf{x}_n^{(r,i)}, \mathbf{Q}, \mathbf{\Omega}^{(i)})$$

- Normalize the weights.

$$w_n^{(r,i)} = w_n^{(r,i)*} / \sum_{r'=1}^{R} w_n^{(r',i)*}$$

M-Step

- M-step is just weighted least squares (if $\mathbf{Y}$ is fully observed).
- Trick: we can simply stack replicate data sets on top of each other.
- Estimate $\boldsymbol{\pi}^{(i+1)}$ by calculating weighted mean and variance.
- Estimate $\boldsymbol{\Omega}^{(i+1)}$ through a BQ-regression.

Starting Values

- Starting from a unit normal distribution produces a slow moving chain.
- Possibly start based on raw scores based on Q-matrix relevant items to get closer to individual ability.
- Still area of active research.

Conclusions and Future Work

- BQ-Regression works and is fully tested for easy cases: (weights only, arbitrary Q-matrix only, missing data only).
- Still needs more testing in the hard case (weights, missing values, and zeros in Q-matrix).
- Importance sampling still needs more work, particularly, starting values.
- Want to test against MCMC algorithm.

Getting the Software

- The source code is available from `http://pluto.coe.fsu.edu/RNetica/RGAutils.html`
- Currently source package only, eventually binary (possible CRAN release).
- Question to `mailto:ralmond@fsu.edu`.

Introduction

- Goal is to find B, b_0 and $S_{\mathbf{YY.X}}$ subject to restriction Q.
- The jth row can be found by regressing Y_j on the X_k values for which $q_{jk} = 1$.
- The sweep operator (Beaton, 1964; Dempster, 1969) will calculate the appropriate coefficients from the covariance matrix.

$$\mathrm{SWP}[k]\mathbf{M} = \begin{bmatrix} m_{ij} - m_{ik}m_{kj}/m_{kk} & m_{ik}/m_{kk} & m_{ij} - m_{ik}m_{kj}/m_{kk} \\ m_{kj}/m_{kk} & -1/m_{kk} & m_{kj}/m_{kk} \\ m_{ij} - m_{ik}m_{kj}/m_{kk} & m_{ik}/m_{kk} & m_{ij} - m_{ik}m_{kj}/m_{kk} \end{bmatrix} . \tag{3}$$

- Sweep operator can be chained to regress out multiple variables.

Calculating the weighted sum of squares

- Let $\mathbf{Y}_+$ be a matrix formed by joining a column of 1's, $\mathbf{Y}$ and $\mathbf{X}$.
- Let $\mathbf{W}$ be a matrix with the weights on the diagonals (and zeros elsewhere).
- Let $\mathbf{T} = \mathbf{Y}_+^T \mathbf{W} \mathbf{Y}_+$

$$\mathbf{T} = \begin{bmatrix} \sum w & \sum w\mathbf{y} & \sum w\mathbf{x} \\ \sum w\mathbf{y} & \sum w\mathbf{y}^T\mathbf{y} & \sum w\mathbf{y}^T\mathbf{x} \\ \sum w\mathbf{x} & \sum w\mathbf{x}^T\mathbf{y} & \sum w\mathbf{x}^T\mathbf{x} \end{bmatrix} \tag{4}$$

$$\text{SWP}[1]\mathbf{T} = \begin{bmatrix} -1/\sum w & \bar{\mathbf{y}} & \bar{\mathbf{x}} \\ \bar{\mathbf{y}} & \mathbf{S}_{\mathbf{yy}} & \mathbf{S}_{\mathbf{yx}} \\ \bar{\mathbf{x}} & \mathbf{S}_{\mathbf{xy}} & \mathbf{S}_{\mathbf{xx}} \end{bmatrix} \tag{5}$$

Regressing out $\mathbf{X}$

- Now sweep out the rows and columns corresponding to the latent variables $\mathbf{X}$.

$$SWP[1,\mathbf{X}]\mathbf{T} = \begin{bmatrix} * & * & \hat{\mathbf{b}}_0 \\ * & \mathbf{S}_{\mathbf{yy}.\mathbf{x}} & \hat{\mathbf{B}} \\ \hat{\mathbf{b}}_0 & \hat{\mathbf{B}}^T & -\mathbf{S}_{\mathbf{xx}}^{-1} \end{bmatrix} . \tag{6}$$

- To get BQ-regression, just sweep out X values which correspond to the 1's in that row of the Q-matrix.
- Can get residual covariance matrix by calculating residuals and then calculating sum of squares.
- More work is needed if any of the $\mathbf{Y}$ values are missing.

EM model for multivariate normal

- Choose initial estimates for regression parameters, $(\mathbf{b}_0^{(0)}, \mathbf{B}^{(0)}, \boldsymbol{\Sigma}_{\mathbf{yy}.\mathbf{x}}^{(0)})$
- Arrange these as an augmented covariance matrix, $\boldsymbol{\Omega}^{(0)}$ by using the reserve sweep operator.
- Note that $\mathbf{T}$ is a sufficient statistic.
- *E-Step* Calculate $\mathbf{T}^{(i+1)} = E[\mathbf{T}|\boldsymbol{\Omega}^{(i)}]$.
- *M-Step* Use a BQ-regression to find $(\mathbf{b}_0^{(i+1)}, \mathbf{B}^{(i+1)}, \boldsymbol{\Sigma}_{\mathbf{yy}.\mathbf{x}}^{(i+1)})$
- Iterate until convergence.

E-step detail

1. Set up a 0 matrix $\mathbf{T}'$ of the same size as $\mathbf{T}$.
2. Make a copy, $\mathbf{Y}_+^{(i)}$ of the augmented data matrix.
3. For each missing data pattern:
    1. Sweep $\mathbf{\Omega}^{(i)}$ to regress the missing values on the others.
    2. Use regression imputation to impute the missing values in $\mathbf{Y}++^{(i)}$.
    3. Let n_p be the sum of the weights of the missing values. Let $S_{\mathbf{y}_{miss}\mathbf{y}_{miss}\cdot\mathbf{y}_{obs}}$ be the residual covariance matrix. Add $n_p S_{\mathbf{y}_{miss}\mathbf{y}_{miss}\cdot\mathbf{y}_{obs}}$ to the corresponding rows and columns of $\mathbf{T}'$.
4. Calculate $\mathbf{T}^{(i+1)} = \mathbf{Y}_+^{(i)T}\mathbf{W}\mathbf{Y}_+^{(i)} + \mathbf{T}'$